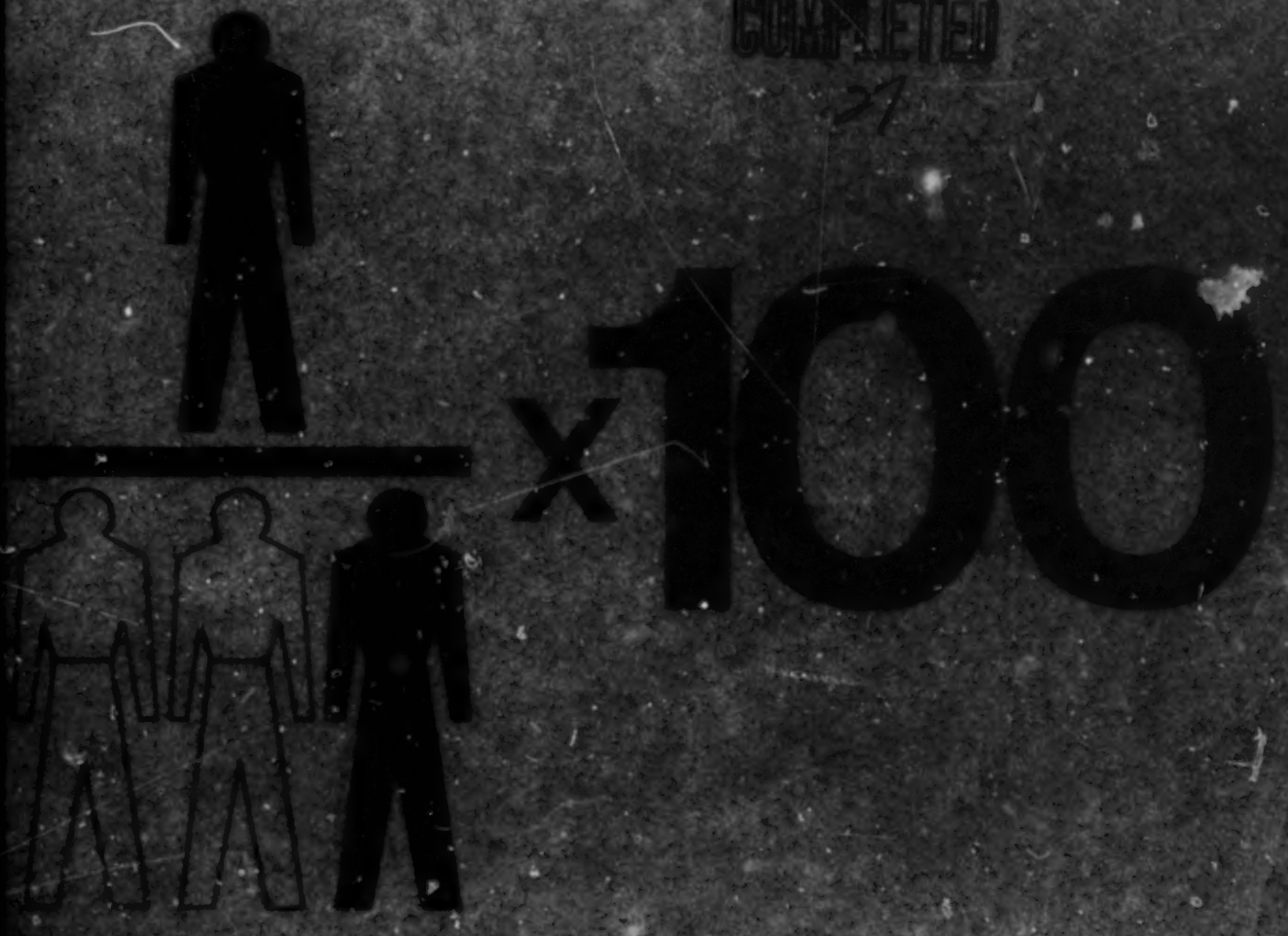
505-A
HE 20.7008:D45
AUG 1 7 1981
ORIGINAL

DESCRIPTIVE STATISTICS

TABLES, GRAPHS, & CHARTS

COMPLETED
-27



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES/PUBLIC HEALTH SERVICE
CENTERS FOR DISEASE CONTROL/ATLANTA, GEORGIA 30333

I

J. Virgil Peavy, M.S.
Director, Instructional Services Division
Center for Professional Development and Training
Centers for Disease Control
Atlanta, Georgia 30333

William W. Dyal
Chief, Health Analysis and Planning for Prevention Services
Center for Prevention Services
Centers for Disease Control
Atlanta, Georgia 30333

Donald L. Eddins
Chief, Statistical Services
Immunization Division
Center for Prevention Services
Centers for Disease Control
Atlanta, Georgia 30333

Reprinted May 1980 Reprinted May 1981

METHOD FOR PRESENTING AND INTERPRETING HEALTH-RELATED DATA

TABLES, GRAPHS, AND CHARTS

I. TABLES

Although there are no hard and fast rules governing table construction, there are certain general principles that have become accepted as more or less standard.

A. Tables should be as simple as possible. Two or three small tables are preferred to a single large table containing many details or variables. Generally, three variables are a maximum number which can be read with ease.

B. Tables should be self-explanatory.

1. Codes, abbreviations, or symbols should be explained in detail in a footnote.
2. Each row and each column should be labeled concisely and clearly.
3. The specific units of measure for the data should be given.
4. The title should be clear, concise, and to the point.
   Answers: what? when? where?
5. Total should be shown.

C. The title is commonly separated from the body of the table by lines or spaces. In small tables, vertical lines separating the columns may not be necessary.

D. If the data are not original, their source should be given in a footnote.

E. Specific examples.

1. The simplest table is a two-column frequency table. The first column lists the classes into which the data are grouped. The second column lists the frequencies for each classification. An example is shown in Table 1.

BEST DOCUMENT AVAILABLE

TABLE 1. CLASSIFICATION OF LIVE BIRTHS
BY EDUCATION OF FATHER, ANYSTATE, 1968

EDUCATION OF FATHER	NUMBER OF LIVE BIRTHS
High School Graduate	50,684
Less than 12 Years of School	31,774
TOTAL	82,458

Source: "Vital Statistics of the United States," 1968, Volume 1, p. 81.

2. Table 1 may be enlarged to include subclassifications, such as place of delivery and attendant at birth, as shown in Table 2.

TABLE 2. CLASSIFICATION OF LIVE BIRTHS BY EDUCATION OF FATHER, PLACE OF DELIVERY, AND ATTENDANT AT BIRTH, ANYSTATE, 1968

Education of Father	NUMBER ATTENDED BY				TOTAL
	Physician		Midwife	Other And Not Specified	
	In Hospital	Not In Hospital			
High School Graduate	46,606	3,014	910	154	50,684
Less Than 12 Years of School	14,334	3,094	13,930	416	31,774
TOTAL	60,940	6,108	14,840	570	82,458

Source: "Vital Statistics of the United States," 1968, Volume 1, p. 81.

3. Summarization of data will be expedited and simplified by initially preparing a master table. In this table, all available data should be completely classified. When complete cross classification is made, data relative to a single variable or to any combination of variables may be obtained without recourse to the original data.

BEST DOCUMENT AVAILABLE

From the general format of a master table as shown in Table 3, we can determine how many URBAN PERSONS (A), how many MALES (B), and how many URBAN MALES IN A SPECIFIED AGE GROUP (C), were admitted to the hospital, etc.

TABLE 3. ADMISSIONS TO ANY HOSPITAL FOR THE YEAR 1968, CLASSIFIED BY AGE, RESIDENCE, AND SEX

AGE IN YEARS	URBAN			RURAL			TOTAL		
	MALE	FEMALE	TOTAL	MALE	FEMALE	TOTAL	MALE	FEMALE	TOTAL
< 1									
1-4									
5-9									
10-14	C								
.									
.									
.									
TOTAL			A				B		

BEST DOCUMENT AVAILABLE

II. GRAPHS

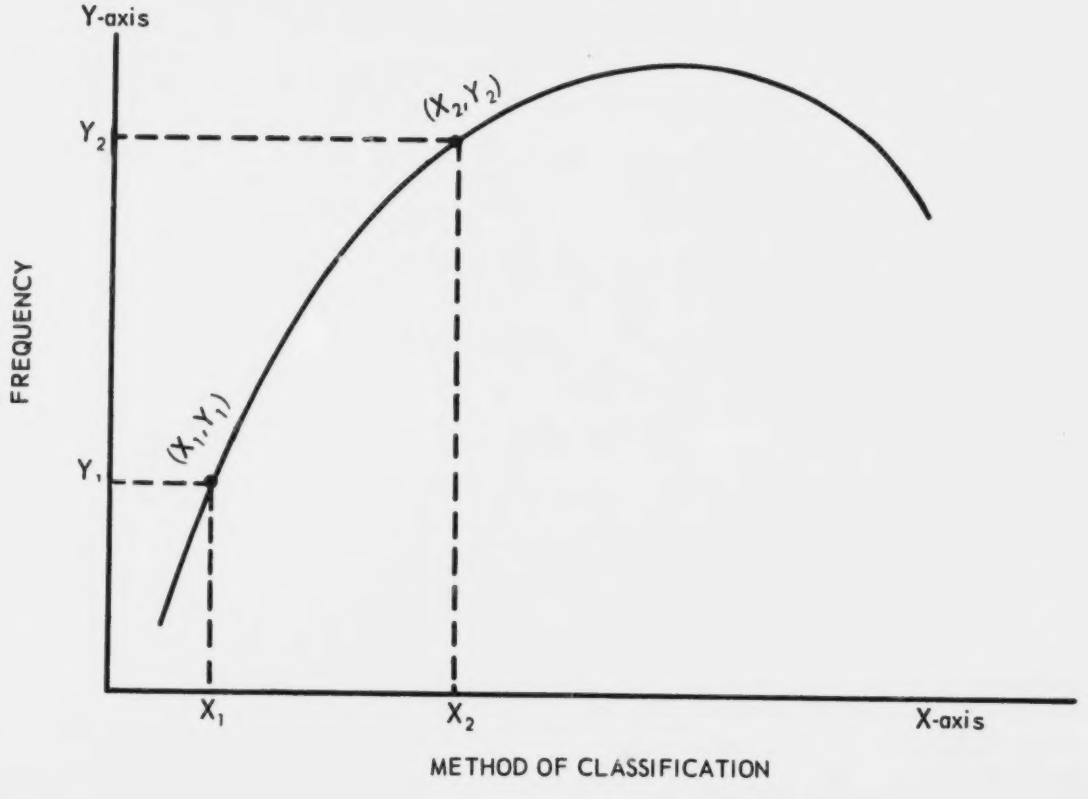
Definition: A graph is a method of showing quantitative data using a coordinate system (for our purposes, usually x and y).

There are several different types of graphs, a few of which are rectangular coordinates, polar coordinates, special purpose types (3 dimensional), etc. We will limit ourselves to rectangular coordinate graphs.

A. General Concept

Rectangular coordinate graphs are those which consist of two sets of lines which are at right angles to each other. On each set of lines there is a scale of measurement for purpose of identification. Figure 1 presents the general structure of the rectangular coordinate graphs. Generally the variable assigned to the x-axis is considered the independent variable (method of classification), whereas the variable assigned to the y-axis is the dependent variable (frequency). That is, in drawing a graph, we plot a change in "y" with respect to x.

Figure 1
GENERAL GRAPH



BEST DOCUMENT AVAILABLE

B. General Principles

When graphs have been drawn correctly, they allow the reader to rapidly obtain an overall grasp of the data. Some of the most important principles of graphing are:

1. The simplest graphs are the most effective. No more lines or symbols should be used in a single graph than the eye can easily follow.

2. Every graph should be self-explanatory.

3. The title may be placed either at the top or bottom of the graph.

4. When more than one variable is shown on a graph, each should be clearly differentiated by means of legends or keys.

5. No more coordinate lines should be shown than are necessary to guide the eye.

6. Lines of the graph itself should be heavier than other coordinate lines.

7. Frequency is usually represented on the vertical scale and method of classification on the horizontal scale.

8. On an arithmetic scale, equal increments on the scale must represent equal numerical units.

9. Scale divisions should be clearly indicated as well as the units into which the scale is divided.

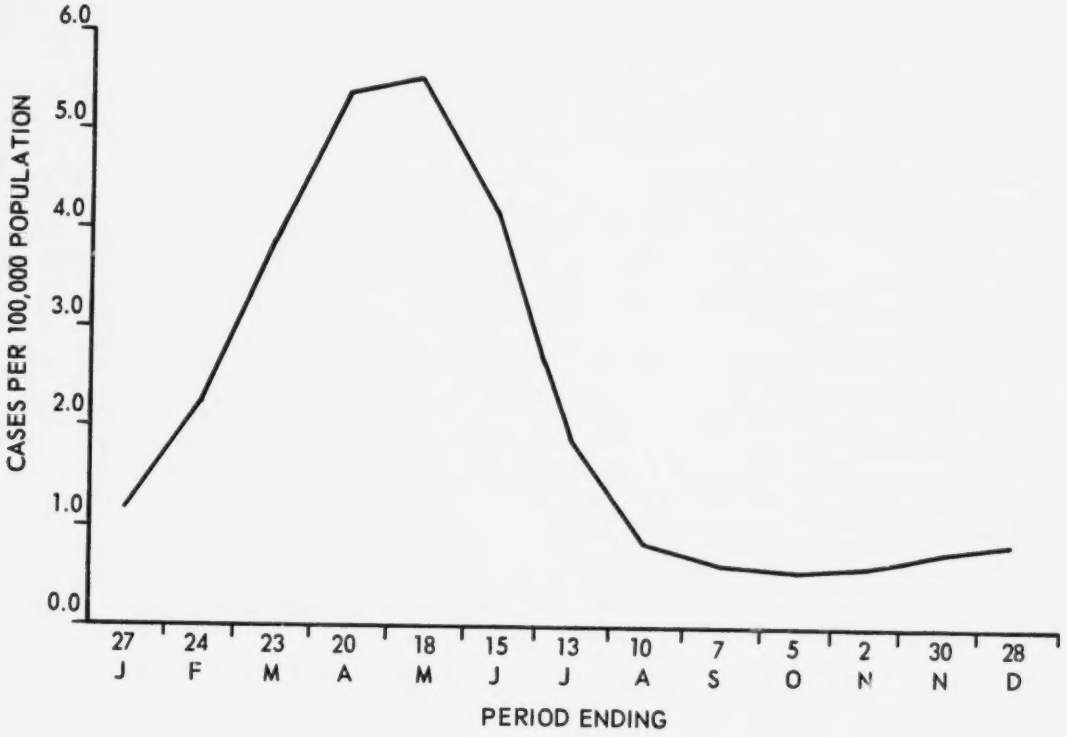
BEST DOCUMENT AVAILABLE

C. Specific Examples

1. Arithmetic Scale Line Graph

A scale line graph is one where an equal distance represents an equal quantity anywhere on the axis, but not necessarily between the axes. Care must be exercised in the choice of whether we use equal intervals on both axes, wide intervals on the x-axis in relation to the y-axis or vice versa. The scales should be defined in such a way that the final product is pleasing to the eye. A scale break may be used with a scale line graph, but if used, care must be taken lest the graph be misinterpreted. Figure 2 is an example of a scale line graph.

Figure 2
REPORTED RUBELLA CASE RATES BY 13 FOUR-WEEK PERIODS,
UNITED STATES 1968



Source: Rubella Surveillance Unit.

BEST DOCUMENT AVAILABLE

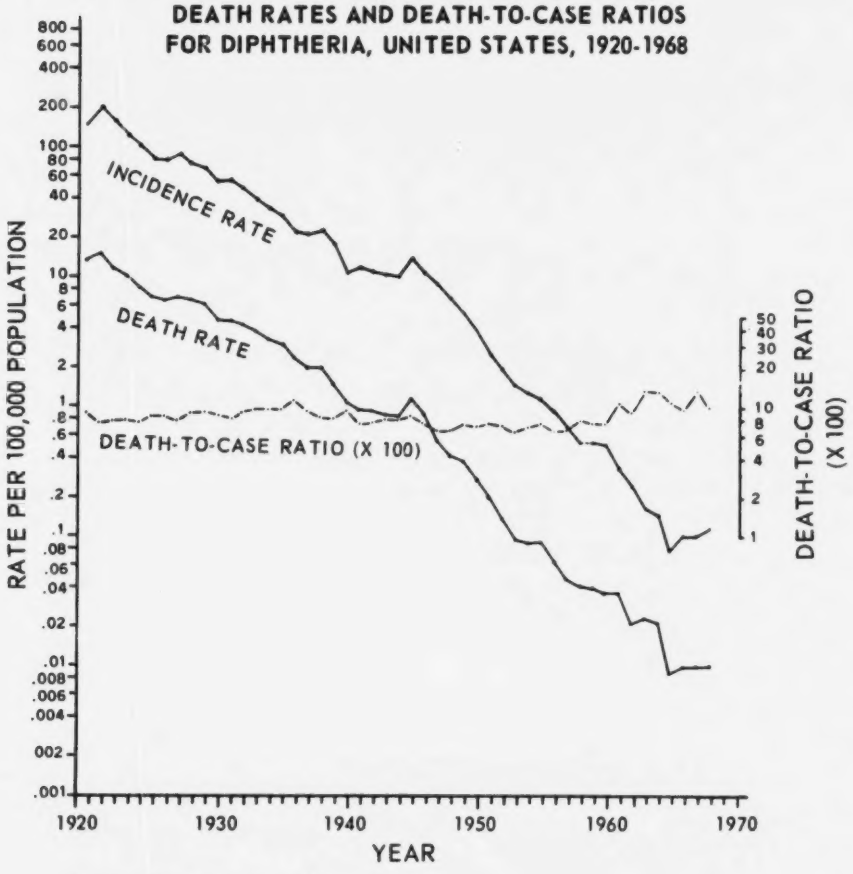
2. Semi-logarithmic Scale Line Graph

The semi-logarithmic scale line graph is one where one coordinate or axis, usually the y-axis, is measured in logarithms of units, whereas the other axis is measured in arithmetic units. This is useful in that when examining a series of data over a period of time we are often interested in the relative (or rate of) change rather than the absolute (actual amount of). The advantages of semilog graphing are:

a) a straight line indicates a constant rate of change,
b) the slope of the line indicates the rate of increase or decrease,
c) two or more lines following parallel paths show identical rates of increase (or decrease).

An illustration of this type of graph is shown in Figure 3.

Figure 3
REPORTED ANNUAL INCIDENCE RATES,
DEATH RATES AND DEATH-TO-CASE RATIOS
FOR DIPHTHERIA, UNITED STATES, 1920-1968



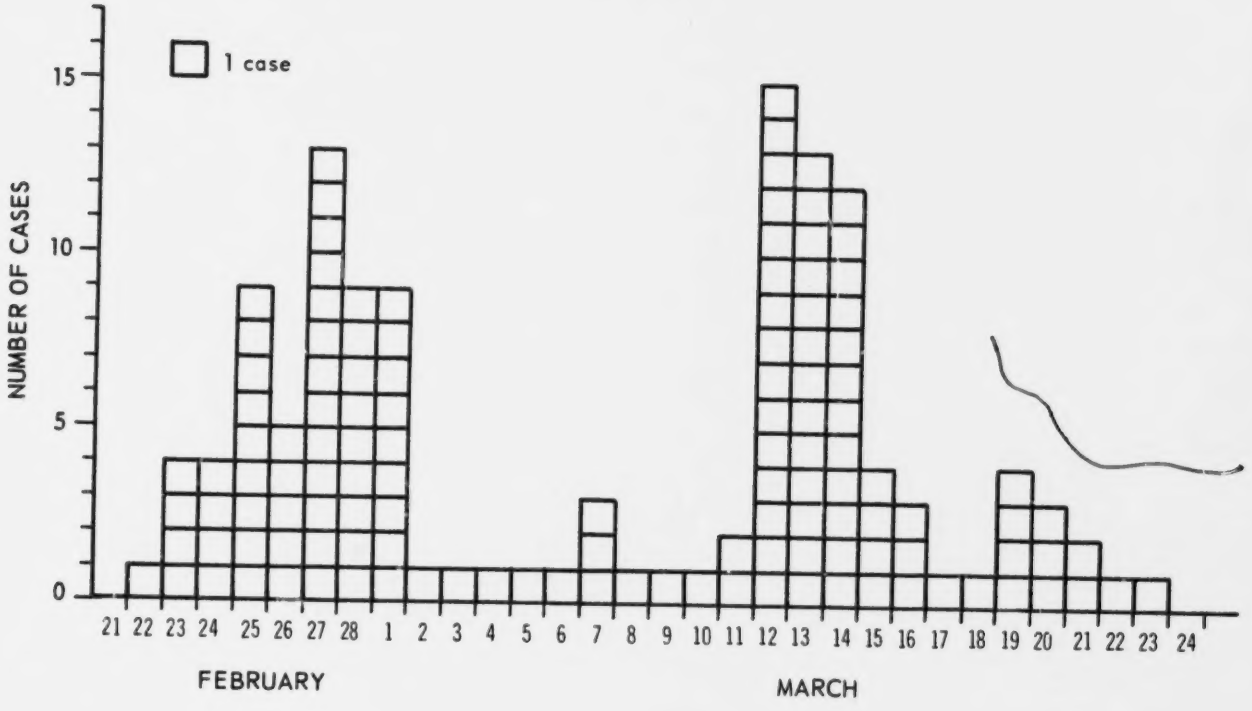
Source: Diphtheria Surveillance Unit.

BEST DOCUMENT AVAILABLE

3. Histogram

A histogram is a graph used only for presenting frequency distribution of quantitative data. There is no space between the cells (often referred to as tic-marks) on a histogram. This graph is not to be confused with a bar chart which has space between the cells. A scale break should not be used in the histogram because the histogram depicts the total area under the curve. Because of this characteristic the easiest type of histogram to construct will be one of equal class intervals as shown in Figure 4.

Figure 4
CASES OF RASH ILLNESS, ELEMENTARY SCHOOL
SAMPLE CITY, FEBRUARY 22 - MARCH 23, 1970



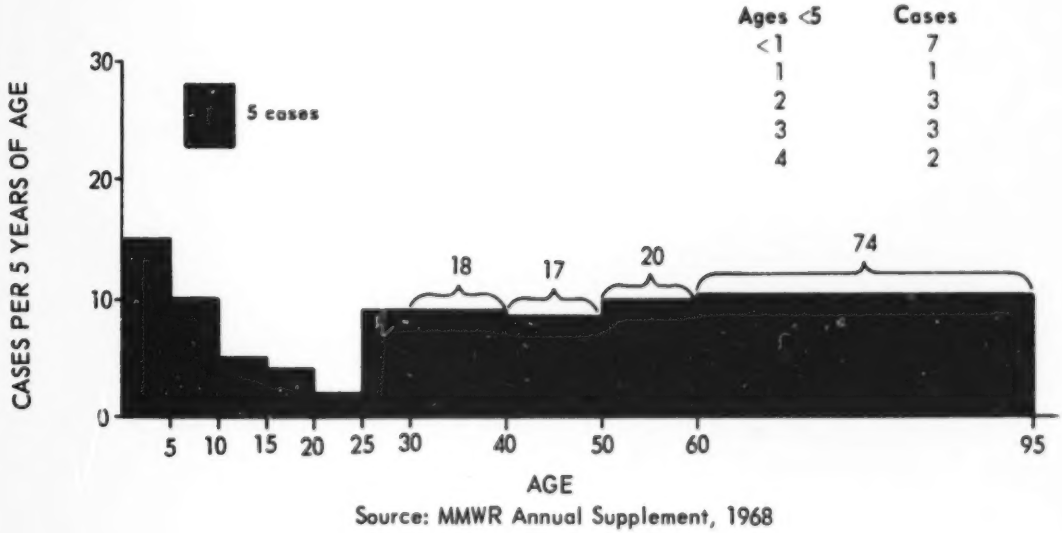
BEST DOCUMENT AVAILABLE

For illustration, Figure 4 shows the area under the curve partitioned into each case of illness. Ordinarily, only the line representing the height of each column would be drawn.

In order that the area of each rectangle in the histogram represents a specified number of cases, we let the heighth represent the number of cases per unit of measurement (in Figure 4, per day) and the width be the method of classification (in Figure 4, interval of time in days). Therefore, the heighth times the width will equal the number of cases within a day, just as heighth times width is equal to the area of a rectangle.

A histogram with unequal intervals requires more thought in its construction because of the total area concept. Figure 5 has been selected to illustrate this situation.

Figure 5
REPORTED CASES OF TETANUS BY FIVE-YEAR AGE GROUPS
UNITED STATES 1968



Source: MMWR Annual Supplement, 1968

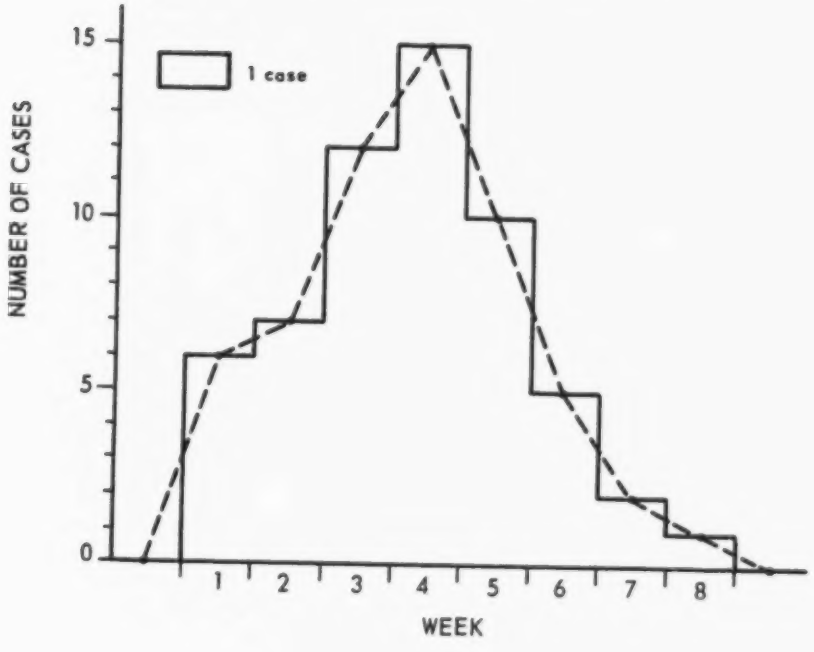
In general, only one set of data should be shown on a histogram; however, it is quite common in the field of public health to present data for cases-deaths, males-females, etc., by the histogram.

BEST DOCUMENT AVAILABLE

4. Frequency Polygon

If it is desired to present more than one set of data in terms of a frequency distribution, the data should be presented in the form of a frequency polygon. A frequency polygon is constructed from a histogram, i.e., the midpoints of the class intervals are connected by a straight line. This is illustrated in Figure 6.

FIGURE 6
NUMBER OF CASES OF INFLUENZA-LIKE ILLNESS BY WEEK, SAMPLE CITY, 1970



BEST DOCUMENT AVAILABLE

Since a frequency polygon is constructed from a histogram, the rules pertaining to area under the curve, equal and unequal class intervals will hold. Therefore, it is necessary to illustrate the manner in which this concept is preserved. The frequency polygon should be "closed" in order to portray area. This "closing" is done by connecting the first and last points with the base of the graph. The area in the frequency polygon must be approximately equal to that which would have been in the histogram. Figure 7 shows the correct method of closing the frequency polygon. Figure 8 shows the incorrect method.

Figure 7
CORRECT METHOD OF CLOSING FREQUENCY POLYGON

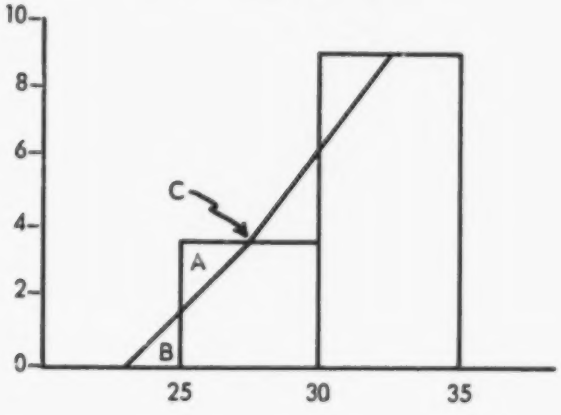
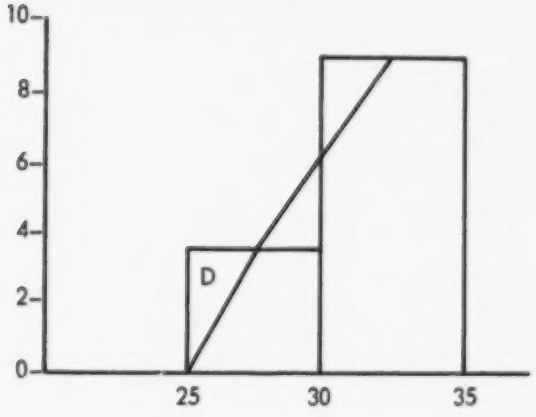


Figure 8
INCORRECT METHOD OF CLOSING FREQUENCY POLYGON



In Figure 7 the area designated by A would be part of the histogram if data were plotted by that method. In order to compensate for this area, which is excluded by the polygon, the point C is connected to the base line in such a manner that the area designated by B will be approximately equal to area A.

Figure 8 illustrates the incorrect method of closing a frequency polygon because the entire area designated by D is omitted and there is no provision for compensation.

BEST DOCUMENT AVAILABLE

In Figure 9 a frequency polygon having equal class intervals is shown. Additionally, Figure 10 illustrates a frequency polygon with unequal class intervals.

Figure 9
NUMBER OF CASES OF INFLUENZA-LIKE ILLNESS BY WEEK, SAMPLE CITY 1970

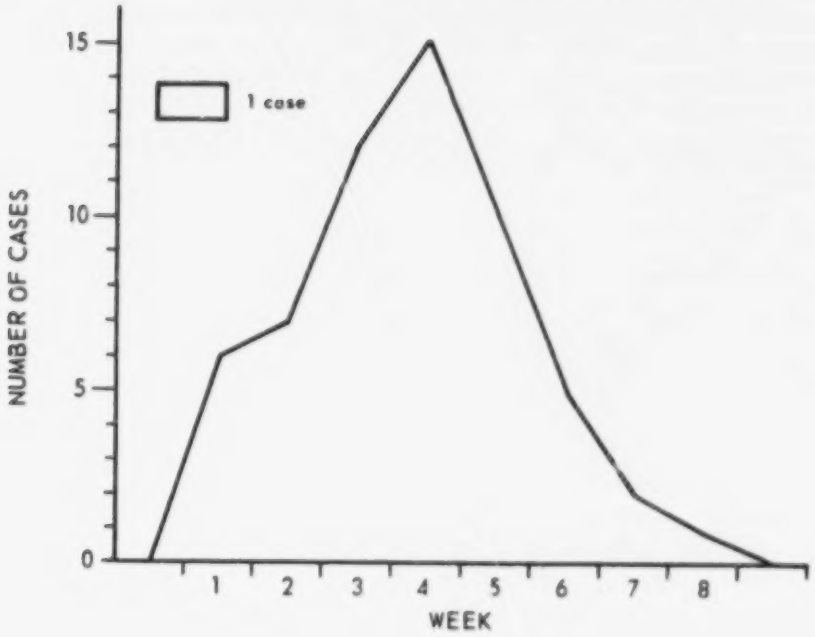
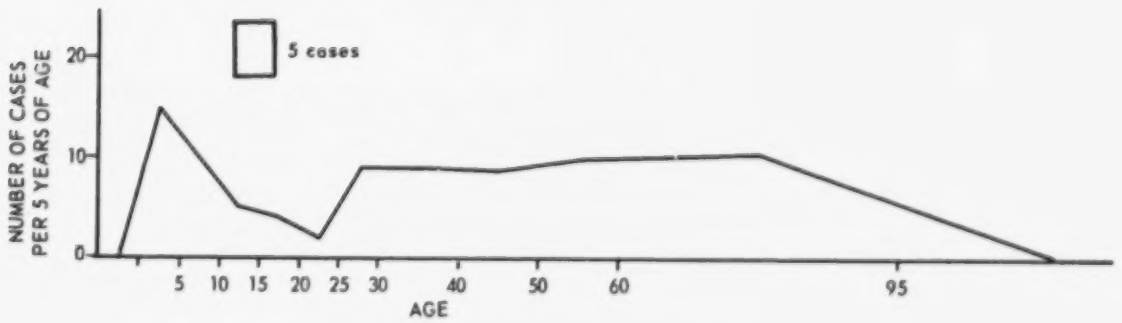


Figure 10
REPORTED CASES OF TETANUS BY FIVE YEAR AGE INTERVALS UNITED STATES, 1968

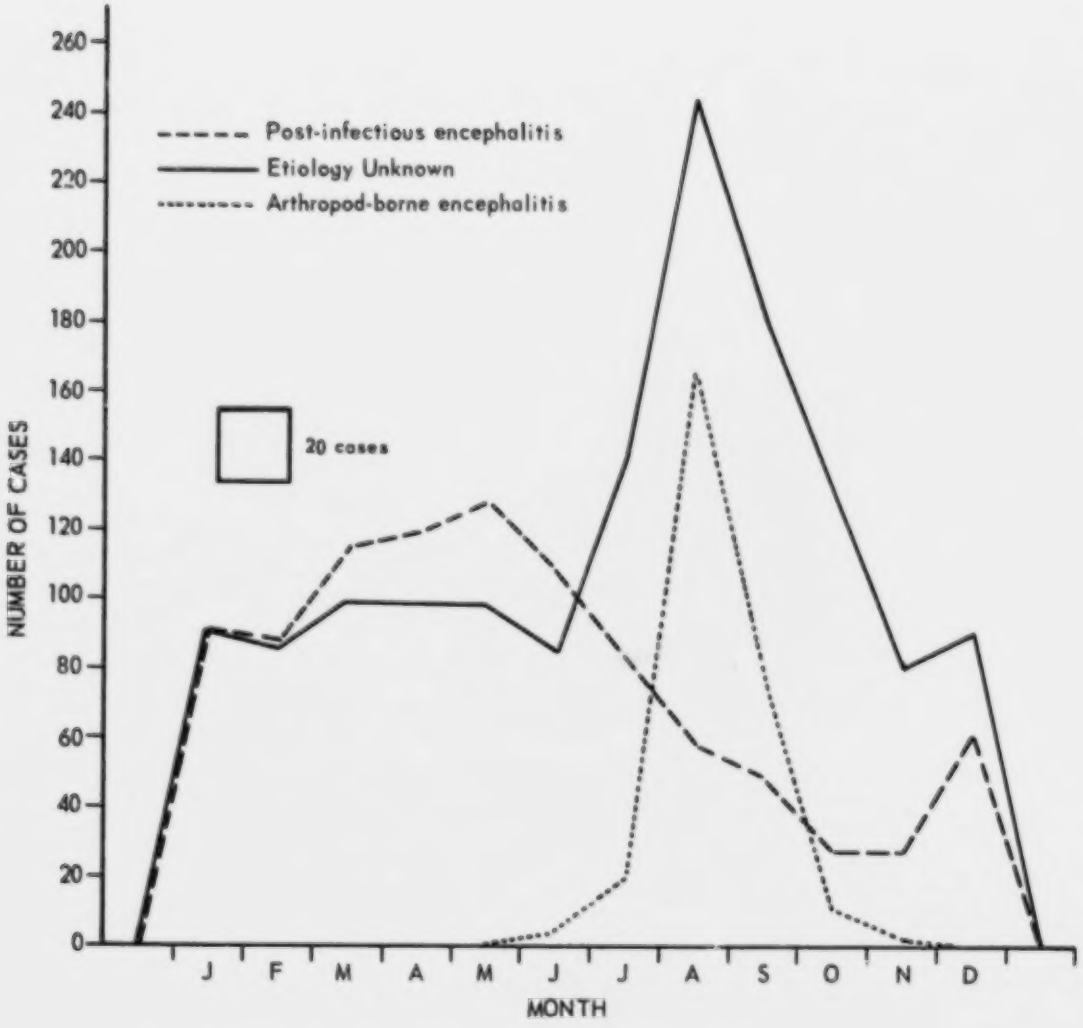


Source: MMWR Annual Supplement, 1968

BEST DOCUMENT AVAILABLE

A frequency polygon illustrating three sets of data is shown in Figure 11.

Figure 11
REPORTED CASES OF ENCEPHALITIS BY MONTH
ACCORDING TO ETIOLOGY, UNITED STATES 1965



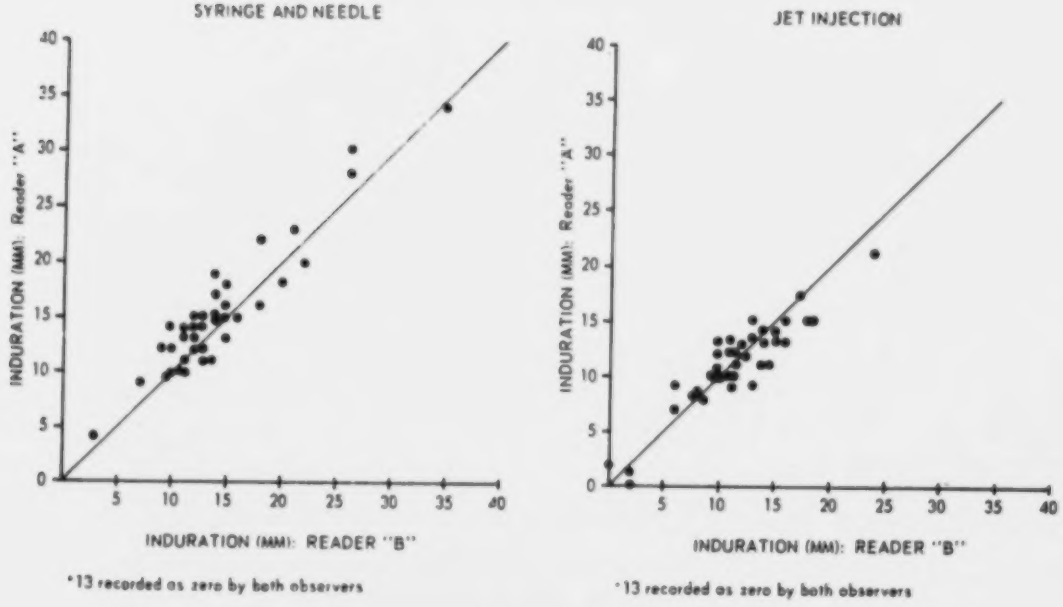
Source: 1965 Annual Summary, Encephalitis Surveillance Report

BEST DOCUMENT AVAILABLE

5. Scatter Diagram

A scatter diagram is a special type of graph useful in pointing out relationships or associations between two variables. In this type of graph several sets of paired data are plotted on the same graph. The pattern made by the plotted points is indicative of a possible relationship. If they tend to follow a straight line, the relationship is of linear nature. If the pattern does not follow a straight line, the relationship is curvilinear. If the pattern is just a scatter of points, then the trend suggests that probably no relationship exists. Figure 12 shows a scatter diagram.

Figure 12
HISTOPLASMIN SKIN TESTS
Comparison of observations of two readers in 51 subjects
Greenwood, South Carolina, 1963



*13 recorded as zero by both observers

*13 recorded as zero by both observers

Source: Miller, J.D., *et al.*: *American Review of Respiratory Disease*, Vol. 100, 1969

BEST DOCUMENT AVAILABLE

III. CHARTS

Definition: Charts are methods of presenting statistical information symbolically using only one coordinate.

There are many different types of charts--some of those based on length, proportion, geographical coordinates, and a few special purpose charts will be presented in this section.

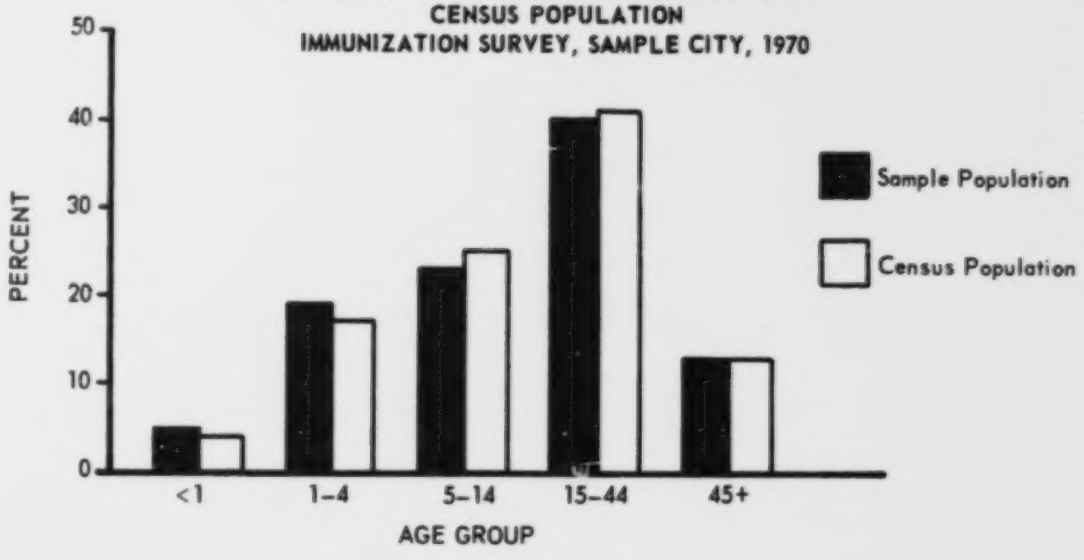
A. Charts Based on Length

Two of the most important charts under this category are bar charts and pictograms.

1. Bar Chart

The bar chart has cells, all of the same column width (unlike the histogram). There are also spaces between the columns (also unlike the histogram). This type of chart is ideally suited for presenting comparative data. The bars may be arranged horizontally as well as vertically (as illustrated in Figures 13 and 14). It is best to arrange these bars in either ascending or descending order for ease of reading. A scale break should never be used with a bar chart as this will lead to misinterpretation. Columns may be shaded, hatched or colored, to emphasize differences between the bars. The bars should be labeled at the bottom and not in the middle of the chart itself, as this adds unnecessary confusion to a chart whose main virtue is its simplicity. When comparisons are made the space between bars in the same group is optional, but space between groups is mandatory.

Figure 13
PERCENT AGE DISTRIBUTION OF SAMPLE POPULATION COMPARED TO PERCENT AGE DISTRIBUTION OF CENSUS POPULATION
IMMUNIZATION SURVEY, SAMPLE CITY, 1970

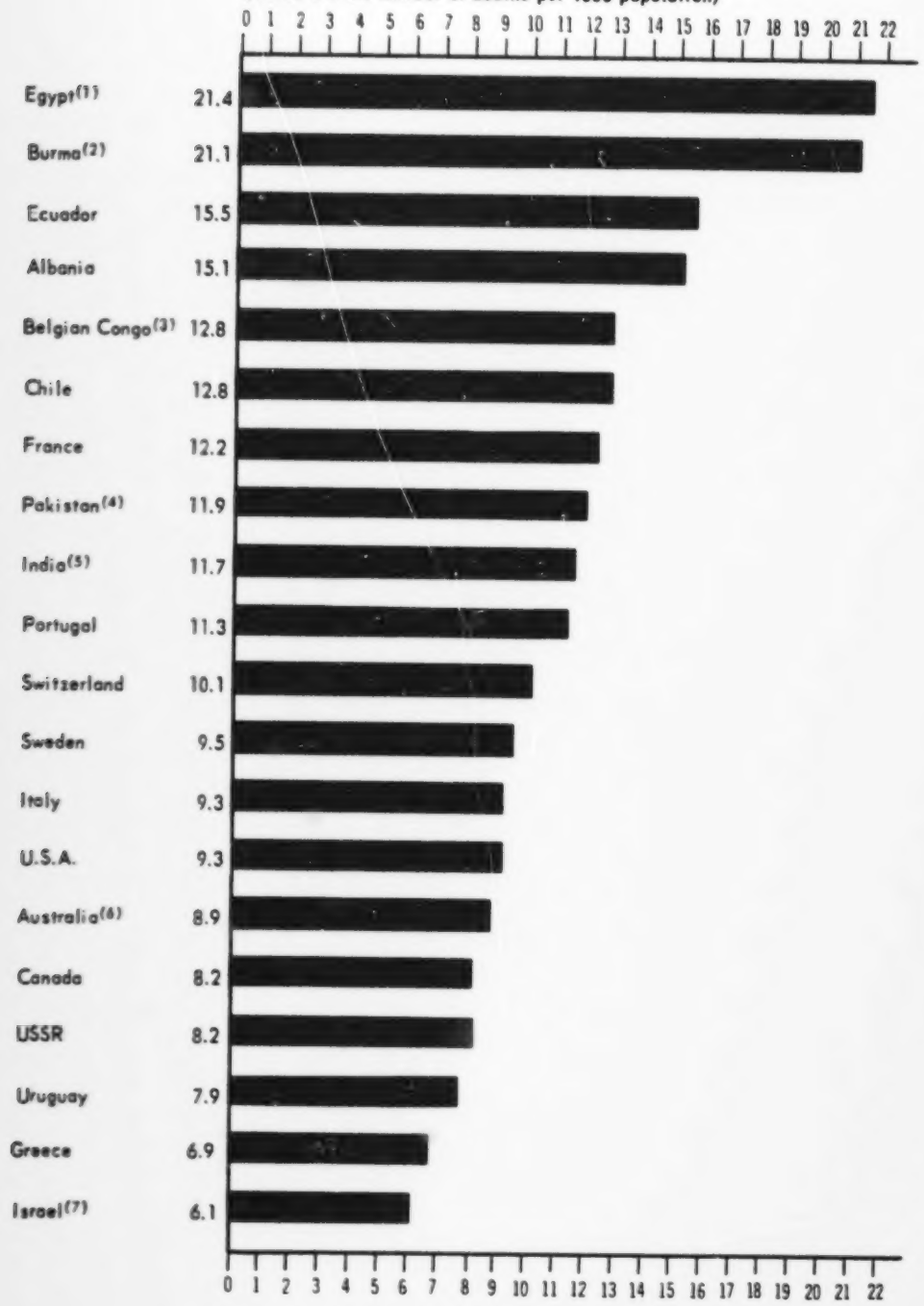


BEST DOCUMENT AVAILABLE

Figure 14

CRUDE DEATH RATE IN 20 COUNTRIES, 1955

(Rates are the number of deaths per 1000 population)


0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Egypt(1) 21.4
Burma(2) 21.1
Ecuador 15.5
Albania 15.1
Belgian Congo(3) 12.8
Chile 12.8
France 12.2
Pakistan(4) 11.9
India(5) 11.7
Portugal 11.3
Switzerland 10.1
Sweden 9.5
Italy 9.3
U.S.A. 9.3
Australia(6) 8.9
Canada 8.2
USSR 8.2
Uruguay 7.9
Greece 6.9
Israel(7) 6.1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

(1) 1954, Health Bureau; (2) 61 towns only; (3) Indigenous population; (4) 1951, Registration area, excluding live-born infants dying before registration of birth; (5) Registration area, excluding data for Bihar, Madhya Pradesh and Orissa; (6) Excluding full-blooded abirugubes; (7) Excluding Bedouin population in the Negev.

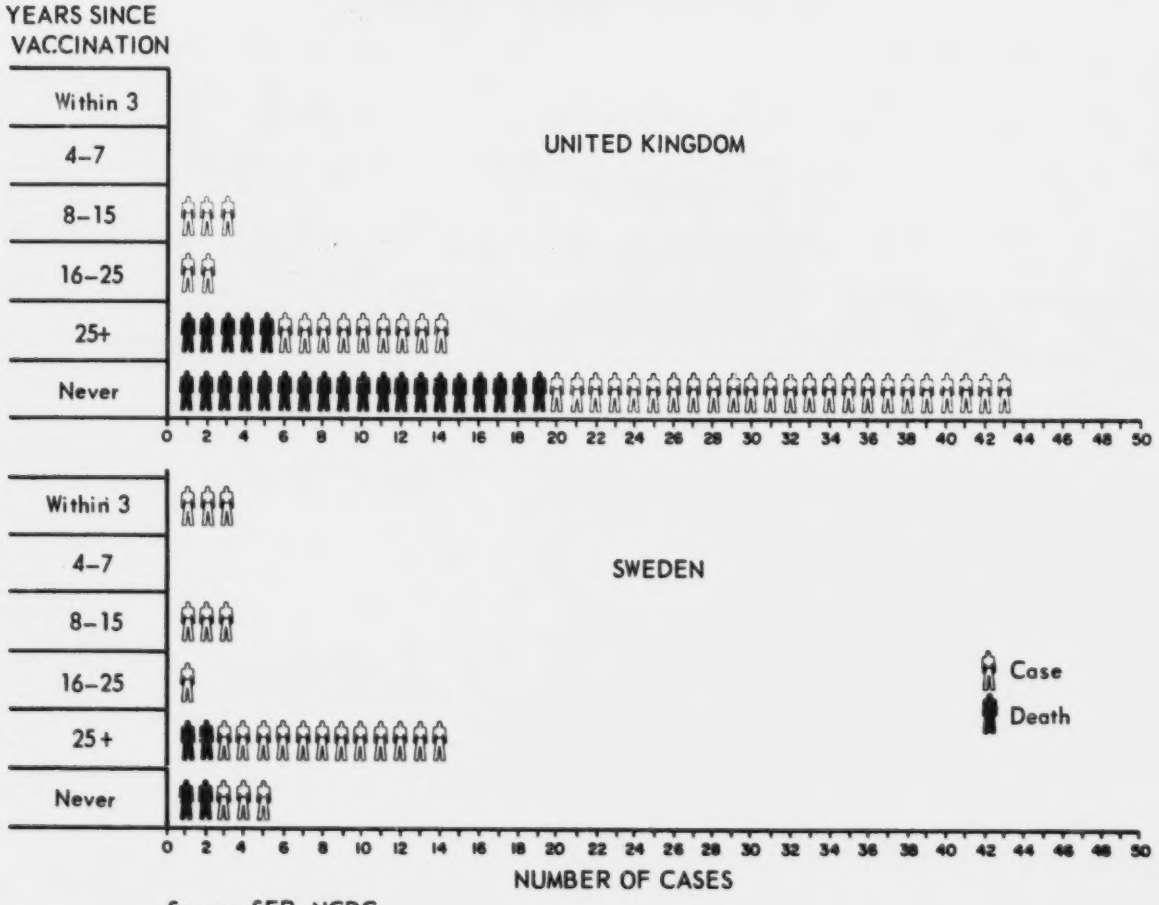
Source: Swaroop, S. *Introduction to Health Statistics*. E. & S. Livingston Ltd., London, 1960

BEST DOCUMENT AVAILABLE

2. Pictogram

The pictogram is a variation on the bar chart using a series of small identifying symbols to present the data. The figures are usually arranged horizontally, but may be arranged vertically. Each symbol may represent a single person or a particular unit. Generally, a symbol will represent a fixed number of persons or units. The number of items is determined by the variable being graphed, which in turn determines the length of the bar. Figure 15 illustrates a pictogram with a single person being represented by a symbol.

Figure 15
VACCINATION STATUS OF SMALLPOX CASES
UNITED KINGDOM–1962, AND SWEDEN–1963



Source: SEP, NCDC

BEST DOCUMENT AVAILABLE

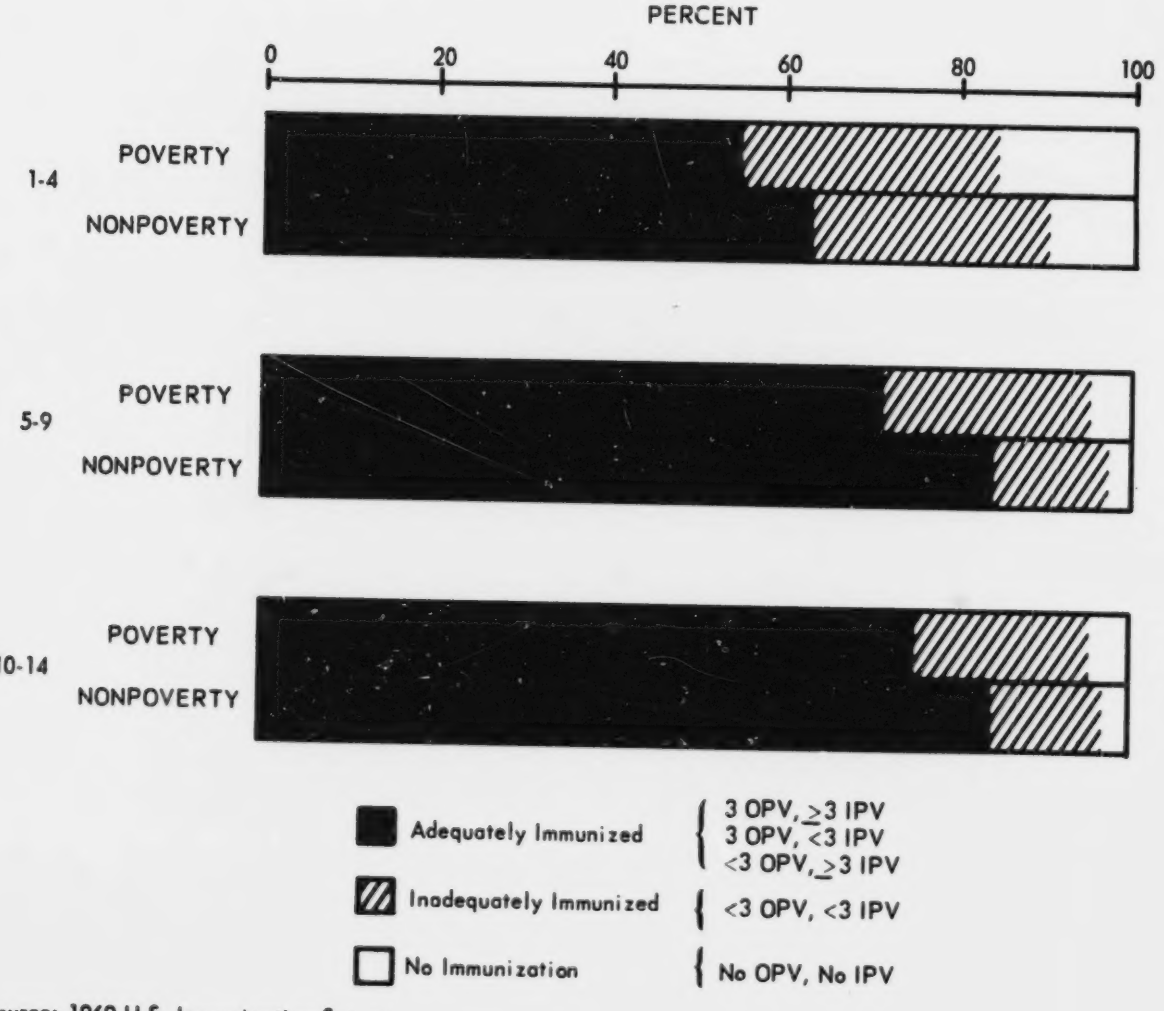
B. Charts Based on Proportion

Two of the most important charts under this category are component bar charts and pie charts.

1. Component Bar Chart

A component bar chart is a bar chart in which the bars are divided into portions which are either colored or shaded to denote their classifications. An example of a component bar chart is given in the following Figure 16.

Figure 16
POLIOMYELITIS IMMUNIZATION STATUS FOR CENTRAL CITIES (POP. ≥250,000) BY AGE AND FINANCIAL STATUS UNITED STATES, 1969



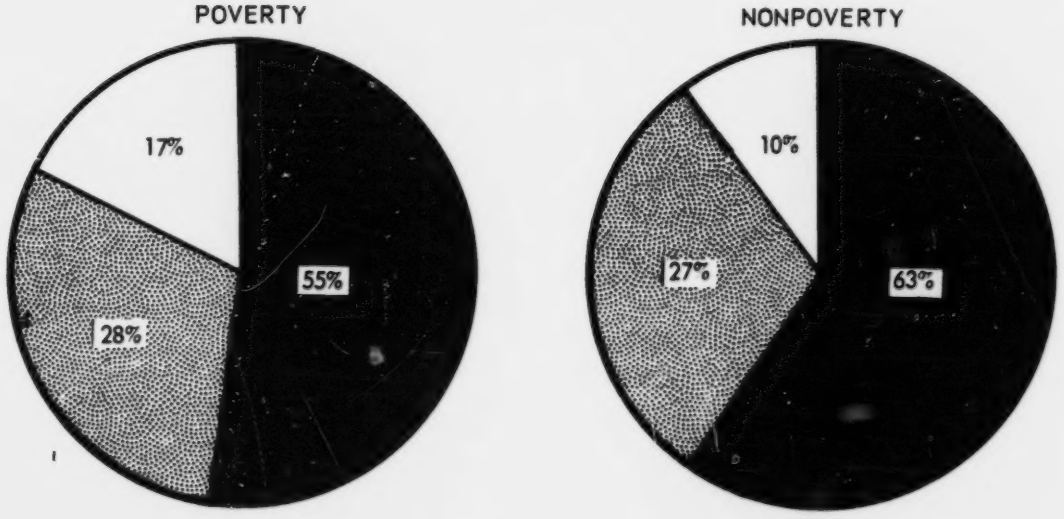
Source: 1969 U.S. Immunization Survey

BEST DOCUMENT AVAILABLE

2. Pie Chart

Pie charts are charts which use wedge shaped portions of a circle for comparison. This chart is sometimes called a sector chart. The pie chart is best adapted for illustrating the division of the whole into segments. The convention is to start at the 12 o'clock position and arrange segments in the order of their magnitude, largest first, and proceed clockwise around the chart. To convert from percentage to degrees, multiply the percentage by 3.6, since 360°/100% = 3.6°%. Figure 17 is an example of a pie chart.

Figure 17
POLIOMYELITIS IMMUNIZATION STATUS OF CHILDREN AGE 1-4 IN CENTRAL CITIES (POP. ≥250,000) BY FINANCIAL STATUS UNITED STATES, 1969



Source: 1969 U.S. Immunization Survey

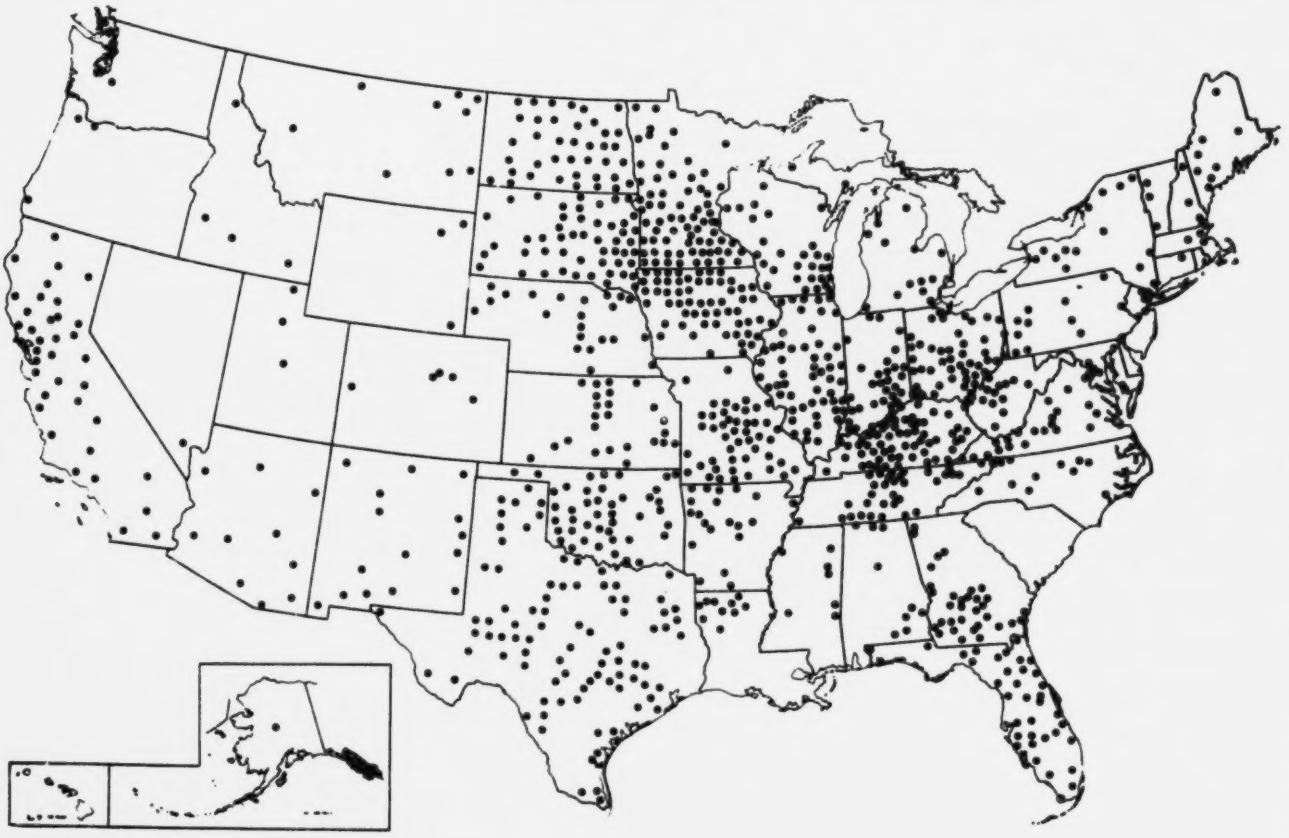
BEST DOCUMENT AVAILABLE

C. Geographic Coordinate Charts

Geographic coordinate charts are those which are based on geographical representation using maps. Disease outbreaks can be very easily plotted on a map for geographic spread. Also, percentages showing immunity levels, etc., can be shown in this fashion. Figure 18 is an illustration of a geographic coordinate chart.

Figure 18

COUNTIES REPORTING ONE OR MORE CASES OF ANIMAL RABIES
UNITED STATES, 1968



• County with one or more cases of animal rabies

Source: MMWR Annual Supplement, 1968.

BEST DOCUMENT AVAILABLE

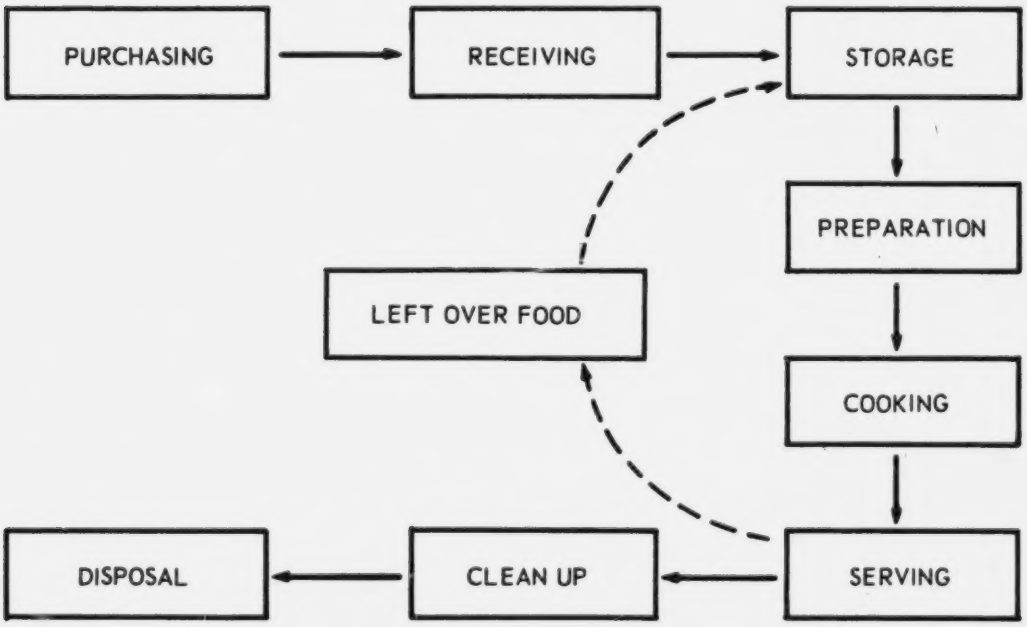
D. Special Purpose Charts

In addition to charts based on length, proportion, and geographical coordinates, other visual methods of presenting data include flow charts, organizational charts, personnel charts, and function charts.

1. Flow Charts

Examples of flow through a sewage treatment plant, flow through a water treatment plant, or such are examples of flow charts. Food flow through a restaurant could be presented visually in a flow chart as in Figure 19.

Figure 19
FOOD FLOW IN JOE'S RESTAURANT, SAMPLE CITY, 1970

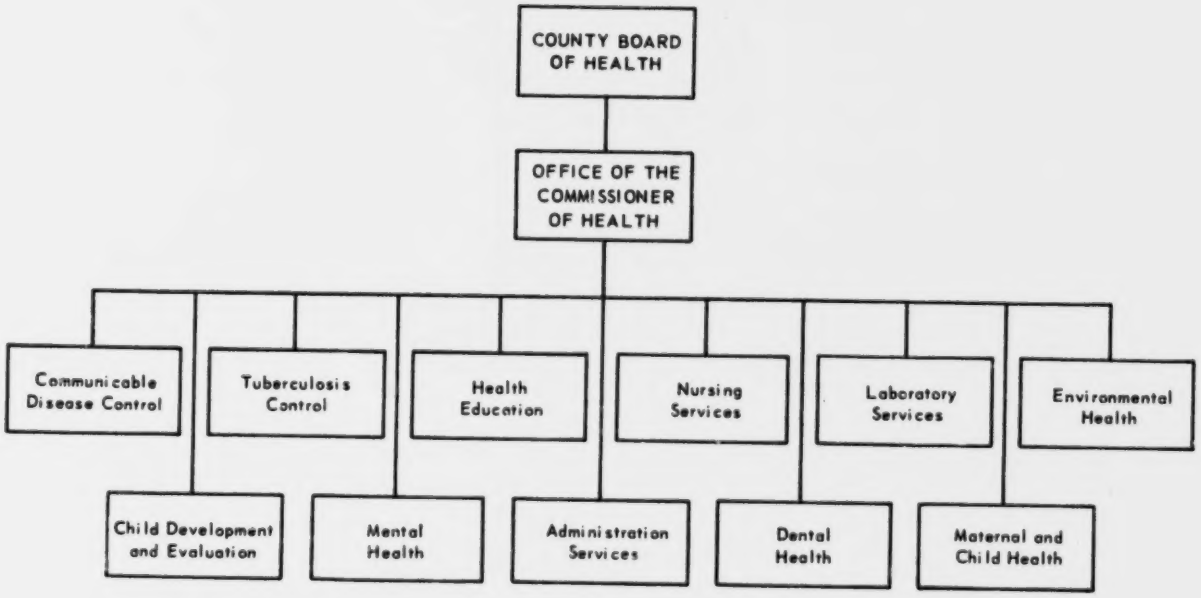


BEST DOCUMENT AVAILABLE

2. Organization Charts

Organization charts are sometimes confused with personnel charts and function charts. Organization charts should show the name of the office, division, or section--not the personnel involved or their function. An example of an organization chart is shown in Figure 20.

Figure 20
ORGANIZATION CHART OF THE TILLER COUNTY HEALTH DEPARTMENT, 1968



BEST DOCUMENT AVAILABLE

IV. SUGGESTIONS FOR THE DESIGN AND USE OF TABLES, GRAPHS, AND CHARTS

In conclusion, it would be well to review and re-enforce what has been discussed up to this point.

A. Choose the Tool most Effective for Data and Purpose

Some methods of presentation call for more complete data than others; a few require special configurations of data. Within such limitations, decide upon the precise idea that you wish to communicate, then choose the method: Continuous line graphs are suitable for a comparison of trends; bar charts clearly compare separate quantities of limited number; pie charts have advantages in comparing parts to their whole; scatter diagrams are excellent for showing tendency.

B. Point Out One Idea at a Time

Confine the presentation to one purpose or idea; limit the amount of data and include only one kind of data in each presentation. Different viewpoints on the information (unless they are being compared) call for separate presentations. So do large quantities of information or various kinds of information.

C. Use Black and White for Exhibits that are to be Reproduced

Few copying machines can reproduce color--and all color reproduction is expensive. Color can be adequately replaced: for areas, by cross-hatching or dotted fields; for lines, by continuous marks, dots, dashes (of different length in different lines), or combinations of the foregoing.

D. Use Adequate, Properly Located Labels

Titles should include the "what, where, and when" that completely identifies the data they introduce. All other labels should be as clear, complete, and easy to understand--and, like the title, they should be outside the frame of the data. Only keys or legends (and these in a neat "box" that sets them apart from the data) should appear within the field of a graph or chart.

E. Give your Sources

Where, or how (or both), the data were obtained, is vital. Verification or further analysis by members of your audience is difficult if not impossible without full disclosure of sources. Also, access to the original information can prove as useful to the audience as either the excerpts that you present or the conclusions that you propose from them.

BEST DOCUMENT AVAILABLE

F. Use Care in Proposing Conclusions

In particular, draw conclusions that reflect the full body of information from which excerpted data were taken; propose only such conclusions as the data that you present can support. But keep in mind that tables, graphs and charts emphasize generalities--at the expense, of course, of detail. Compensate for this distortion, both in design and in comment. Footnote, in a prominent way, any important detail that has been obscured. Avoid conclusions that do not take such distortion into account.

REFERENCES

1. Ipsen, J. & Feigl, P. Bancroft's Introduction to Biostatistics. Harper and Roe, New York, N.Y., 1970.

2. Swaroop, S. Introduction to Health Statistics. E. & S. Livingstone, Ltd., Edinburgh and London, 1960.

3. Schor. S.S. Fundamentals of Biostatistics. G. P. Putnam's Sons, New York, N.Y. 1968.